This is a post-peer-review, pre-copyedit version of an article published in *Self and Identity*

Hypervigilance to Self-Threat:
Further Experimental Evidence for the Mask Model of Narcissism

Mark Hardaker

Roehampton University

Constantine Sedikides

University of Southampton

Elias Tsakanikos

Roehampton University

Corresponding author: Mark Hardaker, Department of Psychology, Roehampton University, Whitelands College, Holybourne Avenue, London, SW15 4JD, UK; e-mail: Mark.Hardaker@roehampton.ac.uk

Abstract

What is it like be a narcissist? According to the mask model, narcissists portray a hard exterior, but possess a soft core. The narcissistic self is fragile. This presumed fragility has been typically operationalized as a discrepancy between explicit and implicit self-esteem, producing inconsistent findings. A reason for the inconclusiveness of over two decades of research may be that narcissism was tested *in situ*. An important exception is work by Horvath and Morf (2009), who obtained support for the mask model under conditions of *self-threat* in sequential priming task followed by a lexical decision task. We report an experiment ($N = 209$) where we test the replicability of their findings with a larger sample and several methodological alterations. In replication, narcissists manifested hypervigilance or defensiveness (i.e., faster reaction times to self-threatening stimuli). However, given ampler time (235 ms as opposed to 149 ms), narcissists switched from defensiveness to self-regulation (i.e., equivalent reaction times to those of non-narcissists). This switch, being rapid and difficult to detect, may explain in part the prior inconclusive findings. Despite transient intrapersonal turbulence in response to self-threat, narcissists quickly regain their composure and re-establish their granite exterior.

Keywords: narcissism, the mask model of narcissism, self-threat, narcissistic fragility, narcissistic defensiveness

Hypervigilance to Self-Threat:
Further Experimental Evidence for the Mask Model of Narcissism

Agentic (or grandiose) narcissism involves a self-lionizing, entitled, vain, and conniving interpersonal orientation in the domain of agency (e.g., competence, intelligence, achievement). Although accumulated evidence over the last two decades has clarified the consequences of narcissists' interpersonal orientation for individuals, groups or organizations, and society (Hermann, Brunell, & Foster, 2018; Roberts, Woodman, & Sedikides, 2018; Sedikides & Campbell, 2017), the fundamental question of "what is it like be a narcissist" remains elusive. This question is the purview of the mask model.

The mask model originates in psychodynamic theorizing (Freud, 1914/1957; Kernberg, 1975; Kohut, 1966). Narcissists, according to this model, may boast a hard exterior, but actually have a soft core. Narcissists are characterized by inner fragility (Kernis, 2003; Westen, 1990). How to operationalize fragility is a challenge. Most typically (but see Mota et al., 2019, for variants), the construct has been operationalized in terms of a discrepancy between explicit self-esteem (as assessed, for example, by the Rosenberg Self-Esteem Scale [RSES]; Rosenberg; 1965) and implicit self-esteem (as assessed, for example, by the self-esteem Implicit Association Test; Greenwald & Farnham, 2000): Narcissists have high explicit, but low implicit, self-esteem.

The bulk of research has tested the mask model assessing narcissism *in situ*. The evidence has been mixed. Some studies obtained support for narcissistic fragility (Gregg & Sedikides, 2010; Jordan, Spencer, Zanna, Hoshino-Browne, & Correll, 2003; Di Pierro, Mattavelli, & Gallucci, 2016; Zeigler-Hill, 2006), but others (Brown & Brunell, 2017; Marissen, Brouwer, Hiemstra, Deen, & Franken, 2016), including an early meta-analysis (Bosson et al., 2008), obtained no support for it even when focusing exclusively on the agentic domain (rather than the communal domain—i.e., warmth, cooperation, relatedness; Campbell, Bosson, Goheen, Lakey, & Kernis, 2007; Fatfouta & Schröder-Abé, 2018). A recent and comprehensive approach, applying an information-theoretic and Response Surface Analysis to data from 18 samples, yielded inconsistent findings (Mota et al., 2019).

A reason for this inconclusiveness may be the conspicuous absence of *self-threat*. The rationale is as follows (Morf, Horvath, &Torchetti, 2011; Morf & Rhodewalt, 2001; Myers & Zeigler-Hill, 2012). Self-threat will perturb the already fragile narcissistic self. Sensing their chronic insecurity, narcissists will be on alert for incoming self-threatening stimuli, detecting them expeditiously. Narcissists' quick responses (i.e., reaction times) to perceived threat, then, is indicative of defensiveness. However, although narcissists will be initially hypervigilant to self-threatening stimuli, they will subsequently self-regulate masking their vigilance and, by correspondence, brittleness. By doing so, they will manage to preserve their steely exterior, a puffed up persona.

Horvath and Morf (2009) tested directly, and supported, the mask model. Their paradigm involved a sequential priming task succeeded by a lexical decision task. The logic behind these tasks follows the theoretical rationale outlined above. Narcissists are assumed to have a deep-seated sense of insecurity or inadequacy (worthlessness). Narcissists are brittle. If so, when they are initially primed with a negative (as opposed to neutral) prime, they will be particularly quick in reacting to words that are likely to expose their insecurity (i.e., worthlessness). Hypervigilance, and faster reaction times, to worthlessness-denoting words, following a negative prime, reflect defensiveness.

Let us describe the paradigm and findings of Horvath and Morf (2009) in more detail. First, participants are exposed subliminally either to a negative prime (i.e., failure) or a neutral prime (i.e., note). Subsequently they decide, as fast as they can, if a string of letters is each a word or a non-word. Some of these letter strings are prototypic of worthlessness (e.g., stupid, incompetent, useless), some are neutral (e.g., glass, diagonal, violet), some are fillers (all negative; e.g., nasty), and some are non-words. As such, the primes can be congruent with the target word (e.g., failure-stupid) or incongruent with it (e.g., failure-glass). It is in the case of prime-word congruence that narcissistic insecurity is most exposed, and it is in this case that reaction times are expected to be faster, as a signature of defensiveness and, by implication, underlying fragility. Critically, the letter strings are presented at two stimulus-onset asynchronies (SOA), short (150 ms) and long (2000 ms). Narcissists demonstrate

heightened responsivity to worthlessness words (but not other types of words) presented after a congruent (vs. incongruent) prime in the short SOA, but not in the long SOA. Narcissists, then, appear to be hypervigilant for self-threat in their social environment (i.e., defensiveness), but quickly suppress their responses—exactly as predicted by this version of the mask model.

The pioneering work of Horvath and Morf (2009) has not been replicated with healthy adults. We aimed to test its replicability with a larger sample (i.e., $N = 209$ vs. $N = 64$ of Horvath & Morf's Study 1). We also made two changes to their experimental paradigm. First, we expanded the pool of stimuli. We used, for example, "humiliation" (rather than "failure") as the negative prime, assuming that the unfavourable connotation of "humiliation" is stronger than that of "failure." Second, and most important, although we kept the duration of the short SOA essentially the same (149 ms to Horvath & Morf's 150 ms), we shortened the duration of the long SOA (235 ms to Horvath & Morf's 2000 ms) to test if narcissists move from defensiveness to self-regulation faster than previously thought. The presence or absence of the operation of limited capacity attention is regulated by the length of SOA. The Horvath and Morf (2009) long 2000 ms SOA gives participants time to engage, focus, and commit limited capacity attention to the word-target however for the shorter 150 ms SOA participants are not given the time to engage focus and commit attention and therefore any observed activation is automatic [automatic activation ensues]. The Horvath and Morf (2009) 150 ms SOA reaction-times results are explained by the facilitation effect of related word-prime to word-target produced by automatic spreading activation (ASA) which occurs without intention or awareness (Collins & Loftus, 1975) Expectancy-based priming and prime repetition arguments aside, we chose to reduce our long SOA because we want to see if the inhibition observed by Horvath and Morf (2009) at 2000msec is present at the later stages of ASA which would show narcissists activate and inhibit worthlessness automatically. That is, we wanted to see if the inhibition shown by narcissists at the Horvath and Morf (2009) long 2000 ms SOA was present at a far shorter SOA which could suggest that inhibition is instinctive and not self-regulation. Relatedness proportion (RP) is the proportion of word-

prime/word-target trials in which the prime and target are semantically related. Our target word prime trial has an RP of .25 and at this RP ASA is thought to start decaying between 167 and 300 msec after prime presentation (Hutchinson Neely & Johnson 2001; Neely, 1977) therefore we chose our longer SOA at the midpoint between these two estimates. Rapid and difficult to detect re-establishment of their intrapsychic equilibrium, following defensiveness, would provide an explanation for why the literature has failed to document the mask model (at least in its classic, psychodynamic version). We note that shorter (< 250) SOAs are robust (Jiang et al., 2016; Perea & Gotor, 1997) and less vulnerable to activation decay (Neely, O'Connor, & Calabrese, 2010), processing delays (Kazanas & Altarriba, 2016), or phasic affective modulation (Topolinski & Deutsch, 2013). Also, priming effects found with longer SOAs are more unstable and open to alternative explanations (Wentura & Degner, 2010). Finally, we controlled for self-esteem given its known positive association with narcissism (Brummelman, Thomaes, & Sedikides, 2016), as did Horvath and Morf.

Method

Participants and Design

We tested 209 University of Roehampton psychology students (85.2% women), ranging in age from 18 to 55 years (M = 21.81, SD = 5.26). The sample comprised 177 undergraduate and 32 graduate students. The basis for determining our sample size was Horvath and Morf (2009, Study 1), who reported an effect size of η_p^2 = .07 (N = 64). Using this as a guide, we conducted a G*Power analysis (f^2 = .075; α = .05; β = .95; 2 predictors), which yielded an N of 209 (Faul, Erdfelder, Lang, & Buchner, 2007). For all experiments, we have reported all measures, conditions, data exclusions, and how we determined our sample sizes, and this data is available at https://osf.io/7rxae/

Novel Features of Our Paradigm

We specify the differences between the current paradigm and that of Horvath and Morf (2009). The *first difference* concerned stimuli. The negative prime was "humiliation" (instead of "failure") and the neutral prime was "note" (as in Horvath and Morf). As we mentioned above, we reasoned that the unfavorable meaning of "humiliation" is more potent

than that of "failure" and instead of being an agentic threat like failure, humiliation is the communal threat of public shame Although we still used 16 "worthlessness" words, we included achievement failures, interpersonal rejections, and private as well as public setbacks (again, in an effort to strengthen the meaning of "worthlessness"). We also used different sets of neutral, filler-negative, and non-words, following validational procedures (see below). The *second difference* concerned SOAs. The short SOA was 149 ms and the long SOA was 235 ms instead of Horvath and Morf's 150 ms and 2000 ms, respectively. (The SOA breakdown was as follows. For the short SOA: prime = 35 ms; second mask = 24 ms; target letter string = 90 ms. For the long SOA: prime = 35; second mask = 24; target letter string =176.)

Procedure

We tested participants individually in an enclosed cubicle. We seated them in front of a 21-inch CRT monitor set at an 85 Hertz refresh rate, and gave them verbal instructions regarding the task along with a 1-min practice trial. Then, we asked them to complete 384 pseudo-randomized test trials, which were divided into two blocks of 192. Each trial began with the presentation of a fixation cross, which remained on the screen for 505 ms. This was immediately followed by (1) a brief flickering of letters that contained the first mask (KQHYTPDQFPBYL) for 153 ms, (2) one of two primes (HUMILIATION or NOTE) for 35 ms, and (3) the second mask (FYVDLTMHQWSPW) for 24 ms. We used sandwich masking to prevent prime afterimages (Draine & Greenwald, 1998). We asked participants to concentrate on the fixation cross, and mentioned (ostensibly) that the flickering of letters was due to the program software randomly selecting either a word or a non-word.

Following the masking procedure, we displayed a blank screen for either 90 ms (resulting in a short SOA of 149 ms) or 176 ms (resulting in a long SOA of 235 ms), and then presented participants with one of the 96 letter strings (Appendix A). (We re-primed participants after presenting them with each letter string.) We instructed them to decide if each letter string was a word or non-word, and to respond by pressing the appropriate button on a response box. The letter strings belonged to one of three categories (16 each): worthlessness (e.g., LOSER, FOOL, INCOMPETENT), neutral (e.g., FOLLOW, LOWER,

USUAL), and filler-negative (e.g., ATTACK, HARM, OFFENSIVE), with the last category aiming to distract participants from the worthlessness adjectives. We selected the worthlessness words from an online thesaurus (Bargh & Chartrand, 2000), selected the neutral and filler-negative words from the Harvard Word Database list of words (Stone, Dunphy, & Smith, 1966), and matched all words for Soundex using the Litscape online database (Bargh & Chartrand, 2000). Further, we created 48 orthographically legal non-words by replacing one letter with a vowel in each word of the worthlessness, neutral, and filler-adjective categories (e.g., LOEER, FOLAOW, AETACK), resulting in an equal number of words and non-words (Perea & Gotor, 1997). E-prime presented all letter strings at random four times, once after each prime (negative, neutral) x for each SOA (short, long) combination. We gave participants a response window of 1500 ms and asked them to respond as speedily and accurately as possible; we did not record reaction times outside the 1500 ms window.

Lastly, and after probing participants for suspicion (none expressed it), we requested completion of two scales. We assessed narcissism with the 40-item Narcissistic Personality Inventory (Raskin & Terry, 1988; $\alpha = .89$, $M = 10.48$, $SD = 5.18$). For each item, participants chose between two statements, a narcissistic (e.g., "I think I am a special person") and a non-narcissistic (e.g., "I am no better or worse than most people") one. We assessed self-esteem with the 10-item RSES (Rosenberg, 1965; $\alpha = .91$, $M = 18.74$, $SD = 5.46$). A sample item is: "I feel that I'm a person of worth, at least on an equal plane with others" (1 = *strongly disagree*, 5 = *strongly agree*).

Results

Replicating prior findings (Brummelman et al., 2016), we obtained a positive relation between narcissism and self-esteem, $r\,(207) = .18$, $p < .009$. (See Table 1 for descriptive statistics and intercorrelations.). The reaction time variables and the difference scores (short SOA) were normally distributed however the difference scores (long SOA) were slightly negatively skewed. after thorough investigation we decided to include all cases because having a fast (or slow) reaction time does not necessarily constitute an error and removing

outliers did not significantly influence the results. We proceeded to calculate difference scores, following the exclusion of wrong answers. In particular, we subtracted mean reaction times on neutral-prime trials from mean reaction times on negative-prime trials. Thus, negative scores reflect faster responding as a function of the negative prime, whereas positive scores reflect slower responding as a function of the negative prime, relative to the neutral prime (see Table 2 for descriptive statistics and Tables 3 and 4 for intercorrelations between narcissism and reaction times).We then conducted a hierarchical regression analysis on the difference scores, with self-esteem entered in the first step and narcissism in the second step. Narcissism predicted reaction times: Higher levels of narcissism were associated with faster reaction times at short (F [1, 207] = 4.14, p = .017; b = -1.88, t [207] = -2.86, p = .005; R^2 = .04), but not at long SOA (F [1, 207] = .43, p = .65). Regression analysis also showed that when not controlling for self-esteem, narcissism predicted reaction times: Higher levels of narcissism were associated with faster reaction times at short (F 1, 207] = 7.66, p = .006; b = -.02, t [207] = -2.77, p = .006; R^2 = .04), but not at long SOA (F [1, 207] = .76, p < .38). For comparison, regression analysis showed that when controlling for target 149 neutral-prime trials, target 149 threat-prime trials predicted narcissism (F [2, 206] = 4.13, p = .017; b = -.02, t [206] = -2.24, p = .026; R^2 = .04) but when controlling for target 235 neutral-prime trials, target 235 threat-prime trials did not predict narcissism (F [1, 206] = .94, p = .39). Regression analysis showed that narcissism did not predict reaction times for negative filler words at short (F [1, 207] = .86, p = .36) or at long SOA (F [1, 207] = .43, p = .51). This result provides further comparison and demonstrates that narcissists associate threat related words specifically with worthlessness. Self-esteem did not predict reaction times: It was unrelated to reacting times speed at both short (F [1, 207] = .08, p = .78) and long SOA (F [1,207] = .03, p = .87). Using the same formula to analyse the non-threat related neutral word category revealed no significant interaction effects (all Fs < .2).; the observed activation effects are unique to narcissism. SUPPLEMENT? - We considered testing specific facets of the NPI and collected data relating to the Ackerman, Witt, Donnellan, Trzenlewski, Robins, and Kashy (2011) two factor solution and the original Raskin and Terry (1988) seven factor

solution. All tested NPI facets showed good internal consistency (α's above .70). We found that none of the Ackerman et al (2011) NPI facets predicted reaction times at short or long SOA (the Ackerman et al. (2011) GE/EE facet(s) is invalidated by Ackerman, Donnellan, and Robins, 2012), however three maladaptive facets of the Raskin and Terry (1988) solution, superiority, exhibitionism and exploitativeness, correlated with reaction-times at short SOA (no facets correlated at long SOA); multiple regression analysis showed the model accounted for 7% of the variance (F [3, 205] = 5.49; p = .001; R^2 = .07); superiority predicted reaction times (b = -9.38; t (209) = -2.93, p = .004), however exhibitionism (t (209) = -.70, p = .49) and exploitativeness (t (209) = -1.21; p = .23) did not.

Discussion

Two decades of research on the mask model have produced inconclusive findings. We identified a key reason for this inconclusiveness: Relevant studies tested narcissism in situ. An important exception, Horvath and Morf's work (2009), did obtain support for the classic, psychodynamically-based mask model. This investigation did so by manipulating self-threat. Here, we tested its replicability with a larger sample. Also, we examined whether their results would withstand the test of time by introducing technical alterations and, importantly, by shortening the long SOA considerably (from 2000 ms to 235 ms).

In our experiment, as in Horvath and Morf (2009), self-threat disturbed the narcissistic self, exposing its presumed fragility. Narcissists exhibited hypervigilance (i.e., speedier reactions to self-threatening stimuli) or defensiveness. However, given sufficient time (i.e., 235 ms as opposed to 149 ms), narcissists switched from defensiveness to self-regulation (i.e., reaction times equivalent in speed to those of non-narcissists). This rapid and difficult to detect switch might provide one explanation for why prior research has found inconsistent evidence for the mask model. Despite some transient intrapsychic turbulence in response to self-threat, narcissists quickly manage to regain their composure and maintain their granite exterior.

Our research, along with that of Horvath and Morf (2009), constitutes direct support for the classic mask model. Indirect support can be gleaned from three other research streams.

First, narcissists display higher variability in daily affect or affect intensity as well as self-esteem, especially in response to dissatisfying (than satisfying) life events that involve achievement (Bogart, Benotsch, & Pavlovic, 2004; Emmons, 1987; Rhodewalt, Madrian, & Cheney, 1988; Zeigler-Hill, 2006; Zeigler-Hill, Myers, & Clark, 2010). Second, narcissists show greater changes in anger, anxiety, hostility, aggression, and self-esteem, especially in response to failure (than success) feedback (Bushman & Baumeister, 1998; Konrath, Bushman, & Campbell, 2006; Rhodewalt & Morf, 1998; Twenge & Campbell, 2003). Finally, narcissists manifest physiological reactivity—as indicated by cortisol and alpha-amylase—to daily emotionally distressing events (Cheng, Tracy, & Miller, 2013), as well as physiological reactivity—as indicated by cardiovascular indices and cortisol levels—to laboratory induced stress (i.e., the Trier Social Stress Test; Edelstein, Yim, & Quas, 2010; Kelsey, Ornduff, McCann, & Reiff, 2001; Sommer, Kirkland, Newman, Estrella, & Andreassi, 2009).

We focused on the classic mask model, on agentic narcissism, and on healthy adults. We did not address variants of the model (Kuchynka & Bosson, 2018; Mota et al., 2019), other types of narcissism (e.g., communal—Gebauer & Sedikides, 2018; vulnerable—Weiss & Miller, 2018), or the break-down of agentic narcissism into the admiration and rivalry components (Back, 2018; Geukes et al., 2017). We acknowledge that relying on a single prime word to represent self-threat is a significant limitation. Further research may try a number of different self-threat prime words or theoretically heighten self-threat by adding pronouns. Finally, we did not examine pathological narcissism (Weiss & Miller, 2018). Future research would do well to extend the current experimental paradigm to these domains. Regardless, and in closing, we emphasize our key point: Narcissistic fragility, however minimal, is best detected under conditions of self-threat.

References

Back, M. D. (2018). The narcissistic admiration and rivalry concepts. In A. D. Herman, A. B. Brunel, & J. D. Foster (Eds.), *Handbook of trait narcissism. Key advances, research methods, and controversies* (pp. 57-67). New York, NY: Springer.

Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 253-285). New York, NY: Cambridge University Press.

Bogart, L. M., Benotsch, E. G., & Pavlovic, J. D. (2004). Feeling superior but threatened: The relation of narcissism to social comparison. *Basic and Applied Social Psychology, 26*, 35-44. doi:10.1207/s15324834basp2601_4

Bosson, J. K., Lakey, C. E., Campbell, W. K., Zeigler-Hill, V., Jordan, C. H., & Kernis, M.H. (2008). Untangling the links between narcissism and self-esteem: A theoretical and empirical review. *Social and Personality Psychology Compass, 2*, 1415-1439. doi:10.1111/j.1751-9004.2008.00089.x

Brown, A. A., & Brunell, A. B. (2017). The "modest mask"? An investigation of vulnerable narcissists' implicit self-esteem. *Personality and Individual Differences, 119*, 160-167. doi:10.1016/j.paid.2017.07.020

Brummelman, E., Thomaes, S., & Sedikides, C. (2016). Separating narcissism from self-esteem. *Current Directions in Psychological Science, 25*, 8-13. doi:10.1177/0963721415619737

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology, 75,* 219-229. doi:10.1037/0022-3514.75.1.219

Campbell, W. K., Bosson, J. K., Goheen, T. W., Lakey, C. E., & Kernis, M. H. (2007). Do narcissists dislike themselves ''deep down inside"? *Psychological Science, 18*, 227-229. doi:10.1111/j.1467-9280.2007.01880.x.

Cheng, J. T., Tracy, J. L., & Miller, G. E. (2013). Are narcissists hardy or vulnerable? The role of narcissism in the production of stress-related biomarkers in response to emotional distress. *Emotion, 13*, 1004-1011. doi:10.1037/a0034410

Di Pierro, R., Mattavelli, S., & Gallucci, M. (2016). Narcissistic traits and explicit self-esteem: The moderating role of implicit self-view. *Frontiers in Psychology, 7*: 1815. doi:10.3389/fpsyg.2016.01815

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*, 286-303. doi:10.1037/0096-3445.127.3.286

Edelstein, R. S., Yim, I. S., & Quas, J. A. (2010). Narcissism predicts heightened cortisol reactivity to a psychosocial stressor in men. *Journal of Research in Personality, 44,* 565-572. doi:10.1016/j.jrp.2010.06.008

Emmons, R. A. (1987). Narcissism: Theory and measurement. *Journal of Personality and Social Psychology, 52*, 11-17.

Fatfouta, R., & Schröder-Abé, M. (2018). Agentic to the core? Facets of narcissism and positive implicit self-views in the agentic domain. *Journal of Research in Personality, 74*, 78-82. doi:10.1016/j.jrp.2018.02.006

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*, 175-191. doi:10.3758/BF03193146

Freud, S. (1957). On narcissism: An introduction. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 14, pp. 67-105). London: Hogarth Press. (Original work published 1914)

Frick, P. J., & Hare, R. D. (2001). *The antisocial process screening device*. Toronto: Multi-Health Systems.

Gebauer, J. E., & Sedikides, C. (2018). Communal narcissism: Theoretical and empirical support. In A. D. Herman, A. B. Brunel, & J. D. Foster (Eds.), *Handbook of trait narcissism*. *Key advances, research methods, and controversies* (pp. 69–78). New York, NY: Springer.

Geukes, K., Nestler, S., Hutteman, R., Dufner, M., Küfner, A. P., Egloff, B., & ... Back, M. D. (2017). Puffed-up but shaky selves: State self-esteem level and variability in narcissists. *Journal of Personality and Social Psychology, 112*, 769-786. doi:10.1037/pspp0000093

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 76*, 1022-1038. doi:10.1037/KK122-3514.79.6.10?2

Gregg, A. P., & Sedikides, C. (2010). Narcissistic fragility: Rethinking its links to explicit and implicit self-esteem. *Self and Identity, 9,* 142-161. doi:10.1080/15298860902815451

Hermann, A. D., Brunell, A. B., & Foster, J. D. (Eds.). (2018). *Handbook of trait narcissism: Key advances, research methods, and controversies*. New York, NY: Springer.

Horvath, S., & Morf, C. C. (2009). Narcissistic defensiveness: Hypervigilance and avoidance of worthlessness. *Journal of Experimental Social Psychology, 45*, 1252-1258. doi:10.1016/j.jesp.2009.07.011

Jiang, Z., Qu, Y., Xiao, Y., Wu, Q., Xia, L., Li, W., & Liu, Y. (2016). Comparison of affective and semantic priming in different SOA. *Cognitive Processing, 17*, 357-375. doi:10.1007/s10339-016-0771-8

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology, 85*, 969-978. doi:10.1037/0022-3514.85.5.969

Kazanas, S. A., & Altarriba, J. (2016). Emotion word type and affective valence priming at a long stimulus onset asynchrony. *Language and Speech, 59*, 339-352. doi:10.1177/0023830915590677

Kelsey, R. M., Ornduff, S. R., McCann, C. M., & Reiff, S. (2001). Psychophysiological characteristics of narcissism during active and passive coping. *Psychophysiology, 38,* 292-303. doi:10.1111/1469-8986.3820292

Kernberg, O. F. (1975). *Borderline conditions and pathological narcissism.* New York, NY: Jason Aronson.

Kernis, M. H. (2003). Toward a conceptualization of optimal self-esteem. *Psychological Inquiry, 14*, 1-26.

Kohut, H. (1966). Forms and transformations of narcissism. *Journal of the American Psychoanalytic Association, 14*, 243-272. doi:10.1177/000306516601400201.

Konrath, S., Bushman, B. J., & Campbell, W. K. (2006). Attenuating the link between threatened egotism and aggression. *Psychological Science, 17,* 995-1001. doi:10.1111/j.1467-9280.2006.01818.x

Kuchynka, S. L., & Bosson, J. L. (2018). The psychodynamic mask model of narcissism: Where is it now? In A. D. Herman, A. B. Brunel, & J. D. Foster (Eds.), *Handbook of trait narcissism. Key advances, research methods, and controversies* (pp. 89-95–78). New York, NY: Springer.

Marissen, M. A. E., Brouwer, M. E., Hiemstra, A. M. F., Deen, M. L., & Franken, I. H. A. (2016). A masked negative self-esteem? Implicit and explicit self-esteem in patients with Narcissistic Personality Disorder. *Psychiatry Research, 242*, 28-33. doi:10.1016/j.psychres.2016.04.070.

Morf, C. C., Horvath, S., & Torchetti, L. (2011). Narcissistic self-enhancement: tales of (successful?) self-portrayal. In M.D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 399-424). New York: Guilford.

Morf, C. C., & Rhodewalt, F. (2001). Unravelling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry, 12*, 177-196.

Mota, S., Humberg, S., Krause, S., Fatfouta, R., Geukes, K., Schröder-Abé, N., & Back, M. D. (2019): Unmasking Narcissus: A competitive test of existing hypotheses on (agentic, antagonistic, neurotic, and communal) narcissism and (explicit and implicit) self-esteem across 18 samples. *Self and Identity*. Advance online publication. doi:10.1080/15298868.2019.1620012

Myers, E. M., & Zeigler-Hill, V. (2012). How much do narcissists really like themselves? Using the bogus pipeline procedure to better understand the self-esteem of narcissists. *Journal of Research in Personality*, *46*, 102-105. doi:10.1016/j.jrp.2011.09.006

Neely, J. H., O'Connor, P. A., & Calabrese, G. (2010). Fast trial pacing in a lexical decision task reveals a decay of automatic semantic activation. *Acta Psychologica, 133*, 127-136. doi:10.1016/j.actpsy.2009.11.001

Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition, 62*, 223-240.

Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*, 890-902.

Rhodewalt, F., Madrian, J. C., & Cheney, S. (1998). Narcissism, self-knowledge organization, and emotional reactivity: The effect of daily experiences on self-esteem and affect. *Personality and Social Psychology Bulletin, 24*, 75-87. doi:10.1177/0146167298241006

Rhodewalt, F., & Morf, C. C. (1998). On self-aggrandizement and anger: A temporal analysis of narcissism and affective reactions to success and failure. *Journal of Personality and Social Psychology, 74*, 672-673.

Roberts, T., Woodman, T., & Sedikides, C. (2018). Pass *me* the ball: Narcissism in performance settings. *International Review of Sport and Exercise Psychology, 11*, 190-213. doi:10.1080/1750984X.2017.1290815

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Sedikides, C., & Campbell, W. K. (2017). Narcissistic force meets systemic resistance: The Energy Clash Model. *Perspectives on Psychological Science, 12*, 400-421. doi:10.1177/1745691617692105

Sommer, K. L., Kirkland, K. L., Newman, S. R., Estrella, P., & Andreassi, J. L. (2009). Narcissism and cardiovascular reactivity to rejection imagery. *Journal of Applied Social Psychology, 39,* 1083-1115. doi:10.1111/j.1559-1816.2009.00473.x

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis*. Oxford, UK: MIT Press.

Thomaes, S., Brummelman, E., & Sedikides, C. (2018). Narcissism: A social-developmental perspective. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE Handbook of Personality and Individual Differences* (pp. 377-396). New York, NY: Sage.

Thomaes, S., Bushman, B. J., de Castro, B. O., Cohen, G. L., & Denissen, J. J. A. (2009). Reducing narcissistic aggression by buttressing self-esteem: An experimental field study. *Psychological Science, 20*, 1536-1542. doi:10.1111/j.1467-9280.2009.02478.x

Topolinski, S., & Deutsch, R. (2013). Phasic affective modulation of semantic priming. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 39*, 414–436. doi:10.1037/a0028879

Twenge, J. M., & Campbell, W. K. (2003). “Isn’t it fun to get the respect that we’re going to deserve?” Narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin, 29,* 261-272. doi:10.1177/0146167202239051

Weiss, B., & Miller, J. D. (2018). Distinguishing between grandiose narcissism, vulnerable narcissism, and narcissistic personality disorder. In A. D. Herman, A. B. Brunel, & J. D. Foster (Eds.), *Handbook of trait narcissism*. *Key advances, research methods, and controversies* (pp. 3-13). New York, NY: Springer.

Wentura, D., & Degner, J. (2010). Automatic evaluation isn’t that crude! Moderation of masked affective priming by type of valence. *Cognition and Emotion, 24*, 609-628. doi:10.1080/02699930902854587

Westen, D. (1990). The relations among narcissism, egocentrism, self-concept, and self-esteem: Experimental, clinical, and theoretical considerations. *Psychoanalysis and Contemporary Thought, 13,* 183-239.

Zeigler-Hill, V. (2006). Discrepancies between implicit and explicit self-esteem: Implications for narcissism and self-esteem instability. *Journal of Personality, 74*, 119-144. doi:10.1111/j.1467-6494.2005.00371.x.

Zeigler-Hill, V., Myers, E. M., & Clark, C. B. (2010). Narcissism and self-esteem reactivity: The role of negative achievement events. *Journal of Research in Personality*, *44*, 285-292. doi:10.1016/j.jrp.2010.02.005

Table 1. *Intercorrelations between NPI and RSES, and Mean Difference Scores among Worthlessness Words, as well as Means and Standard Deviations*

Measure	*1*	*2*	*3*	*M*	*SD*
1. NPI	-			10.84	5.18
2. RSES	.18*	-		18.74	5.46
3. 149 SOA (ms)	-.20*	.02	-	5.41	48.97
4. 235 SOA (ms)	-.06	.01	.08	.45	52.53

Note: NPI = Narcissistic Personality Inventory; RSES = Rosenberg Self-Esteem Scale; *$p <$ .05.

Table 2

Reaction Time Means in ms (Standard Deviations) as a Function of Prime, Word, and SOA

	Negative Prime		Neutral Prime	
Word	149 SOA (SD)	235 SOA (SD)	149 SOA(SD)	235 SOA (SD)
Worthlessness	566.83 (78.14)	558.85 (80.83)	561.64 (77.67)	558.40 (78.02)
Neutral	599.16 (86.85)	591.79 (87.35)	588.92 (87.87)	585.5 (85.33)
Filler-Negative	598.83 (80.70)	586.17 (82.55)	585.05 (88.54)	581.64 (81.04)
Non-Word	621.19 (85.7)	613.62 (80.47)	611.28 (82.95)	608.52 (83.39)

Table 3 *Intercorrelations Between Narcissism and Reaction Times in the Target/Prime Negative/Neutral/Target/Word and Neutral/Prime Negative/Neutral/Target/Word Categories at Short SOA*

		1	2	3	4	5	6
1	NPI	-					
2	H_149_Neg	.04					
3	H_149_Neu	.06	.86**				
4	H_149_Tar	.01	.85**	.85**			
5	N_149_Neg	.08	.77**	.76**	.74**		
6	N_149_Neu	.10	.76**	.77**	.79**	.79**	
7	N_149_Tar	.12	.77**	.78**	.81**	.81**	.86**

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4 *Intercorrelations Between Narcissism and Reaction Times in the Target/Prime Negative/Neutral/Target/Word and Neutral/Prime Negative/Neutral/Target/Word Categories at Long SOA*

		1	2	3	4	5	6
1	NPI	-					
2	H_235_Neg	.09					
3	H_235_Neu	.05	.86**				
4	H_235_Tar	.05	.82**	.80**			
5	N_235_Neg	.12	.78**	.78**	.78**		
6	N_235_Neu	.08	.77**	.75**	.75**	.84**	
7	N_235_Tar	.09	.71**	.73**	.78**	.86**	.83**

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5 *Intercorrelations Between Difference Scores at Short SOA and Ackerman et al. (2012) and Raskin and Terry (1988) facets of the NPI*

		1	2	3	4	5	6	7	8	9
1	diff_score_shortSOA	-								
2	LA	-.01								
3	GE_EE	.00	.27**							
4	auth	-.09	.87**	.21**						
5	ssf	-.08	.26**	.02	.18**					
6	supr	-.25**	.15*	.33**	.18*	.19**				
7	exhb	-.15**	.29**	.63**	.29**	.03	.32**			
8	expl	-.16**	.38**	.28**	.38**	.32**	.26**	.34**		
9	vnty	.01	.05	.57**	.04	.07	.21**	.31**	.18**	
10	entmt	-.06	.33**	.43**	.31**	.04	.12	.34**	.31**	.00

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix A

Words Used in the Experiment

Worthlessness Words	**Neutral Words**	**Filler-Negative Words**	**Non-Words**
AMATEUR	BOFFIN	ABRASIVE	ABRAEIVE
CHEAT	CHEETAH	ATTACK	AETACK
FAKE	FLASK	BRAZEN	BRAIEN
FALSE	FLUFFY	CRUSH	DEOASTATE
FLOP	FOLLOW	DEVASTATE	EAIL
FOOL	FOX	EXPLOSIVE	ERUSH
INCAPABLE	INCONCLUSIVE	FIRE	FIEE
INCOMPETENT	INCUBATION	GUERRILLA	GUERRILLO
INEPT	INEVITABLE	HARM	HAAM
LIAR	LIQUOR	INHIBIT	INHIOIT
LOSER	LOWER	INTERFERE	INIERFERE
PITIFUL	PITFALL	JAIL	NEUARALISE
STUPID	STOPPED	NEUTRALISE	OFFENSIOE
UGLY	UNSECURE	OFFENSIVE	PROUIBIT
UNSUCCESSFUL	USUAL	PROHIBIT	REAULSE
WORTHLESS	WORDLESS	REPULSE	UXPLOSIVE
			BOAFIN
			CHEEOAH
			UOX
			FEASK
			FLUFFE
			FOLAOW
			INOUBATION
			INCONCLUSIOE
			INEAITABLE
			LOEER
			LIEUOR
			AITFALL
			OTOPPED
			USEAL
			ONSECURE
			WORALESS
			AEATEUR
			EHEAT
			FAOE
			EALSE
			FLEP
			FOOB
			ANCAPABLE
			INCOOPETENT
			ENEPT
			LAAR
			EOSER
			OITFALL
			STAUPID
			UGOY
			UNSUCCESSFAL
			WOROLESS